School of Computing and Information Systems
The University of Melbourne
COMP90049
Knowledge Technologies (Semester 1, 2019)
Workshop exercises: Week 5

Suppose that we have observed the token `lended`, and we have a dictionary as follows:

```
addendum
blenders
commodity
deaden
end
leader
leant
lent
lemonade
pleading
```

1. Assuming that the "correct" (intended) dictionary entry was `lent`, calculate the precision of the following methods of finding approximate entries from the dictionary.
   (a) Neighbourhood search, with a neighbourhood of 1
   (b) Neighbourhood search, with a neighbourhood of 2
   (c) Neighbourhood search, with a neighbourhood of 3
   (d) Global Edit Distance, with a parameter $[m, i, d, r] = [1, -1, -1, -1]$
   (e) Local Edit Distance, with a parameter $[m, i, d, r] = [1, -1, -1, -1]$
   (f) N-gram Distance, where n is 2 (without padding with terminals)
   (g) Using the Soundex transformation, and then looking for exact matches
   (h) Using the Soundex transformation, and then permitting a 1-neighbourhood

---

2. What is the difference between "data retrieval" and "information retrieval"? Why is the latter a knowledge task, but the former is not?
3. (Extension) How many books are there in an average library? How many words are there in an average library? How many documents are there on the World Wide Web? How many words?
4. Identify some different types of "informational needs."
   (a) Give examples of queries which might indicate a particular type of informational need. Are some of them ambiguous for the type of need?
   (b) (Extension) Input some queries of different types of informational needs into a web search engine like Google. Are search engines better at responding to some types of informational need than others? Is there any indication that the search engine is identifying the type of need and tailoring the results toward that?